Generative Transformer Chatbots for Mental Health Support: A Study on Depression and Anxiety

Jordan J. Bird
Department of Computer Science, Nottingham Trent University
Nottingham, United Kingdom
jordan.bird@ntu.ac.uk

Ahmad Lotfi
Department of Computer Science, Nottingham Trent University
Nottingham, United Kingdom
ahmad.lotfi@ntu.ac.uk

ABSTRACT

Mental health is a critical issue worldwide and effective treatments are available. However, incidence of social stigma prevents many from seeking the support they need. Given the rapid developments in the field of large-language models, this study explores the potential of chatbots to support people experiencing depression and anxiety. The focus of this research is on the engineering aspect of building chatbots, and through topology optimisation find an effective hyperparameter set that can predict tokens with 88.65% accuracy and with a performance of 96.49% and 97.88% regarding the correct token appearing in the top 5 and 10 predictions. Examples of how optimised chatbots can effectively answer questions surrounding mental health are provided, generalising information from verified online sources. The results of this study demonstrate the potential of chatbots to provide accessible and anonymous support to individuals who may otherwise be deterred by the stigma associated with seeking help for mental health issues. However, the limitations and challenges of using chatbots for mental health support must also be acknowledged, and future work is suggested to fully understand the potential and limitations of chatbots and to ensure that they are developed and deployed ethically and responsibly.

CCS CONCEPTS

• **Information systems** → Information retrieval; • **Theory of computation** → **Design and analysis of algorithms**; • **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

Chatbots, Natural Language Processing, Transformers, Mental Health

ACM Reference Format:
Jordan J. Bird and Ahmad Lotfi. 2023. Generative Transformer Chatbots for Mental Health Support: A Study on Depression and Anxiety. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '23), July 5–7, 2023, Corfu, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3594806.3596520

PETRA '23, July 5–7, 2023, Corfu, Greece

ACM ISBN 979-8-4007-0069-9/23/07…$15.00
https://doi.org/10.1145/3594806.3596520

1 INTRODUCTION

Mental health is a critical issue that affects millions of people around the world. According to the World Health Organisation (WHO), an estimated 5% of all adults suffer from depression [WHO, 2021]. The WHO also note that, although effective treatment is available, 75% of those categorised as low- and middle-income do not receive treatment. Indeed, awareness and acceptance of poor mental health have steadily improved [Frank and Glied, 2006, Jones and Wessely, 2005], but there is still a significant stigma about the need for professional help [Sickel et al., 2014]. Mental health stigma can act as a barrier for people experiencing depression, anxiety, or other mental health challenges, preventing them from accessing the support they need. The prevalence of mental health stigma has led many people to view online alternatives favourably over physical human interaction [Hanley and Wyatt, 2021].

This knowledge leads to the concept of the online chatbot. In recent years, advances in Natural Language Processing (NLP) have led to the development of chatbots as a tool for promoting mental well-being. Chatbots are computer programs that can simulate a natural conversation, providing support through textual input and output. Given their accessibility and anonymity, they have the potential to help alleviate the stigma associated with seeking help for mental health issues [Abd-Alrazaq et al., 2019].

This paper focuses on the engineering aspect of chatbots for mental health support, with a specific focus on answering questions about depression and anxiety. The study will explore hyperparameter space to build chatbots based on attention mechanisms and transformers, which are large language models. These models have shown great success in various natural language processing tasks and have the potential to provide effective and engaging support to individuals experiencing mental health challenges. Furthermore, the paper will present examples of interactions with optimised chatbots to demonstrate their effectiveness and usability. The main goal of this work is to contribute to ongoing research in the field of mental health and technology by exploring the potential of chatbots to provide accessible and effective support for people experiencing depression and anxiety.

The remaining parts of this paper are organised as follows; background and related work is presented in Section 2 followed by the proposed method in Section 3. The results and observations are presented in Section 4. Section 5 presents the conclusion and future work.

2 BACKGROUND AND RELATED WORK

Chatbots are Human-Computer Interaction (HCI) models that allow users to converse with machines through natural language [Bansal and Khan, 2018]. Most often in the modern literature, chatbots make use of artificial intelligence and machine learning to process an input and produce a response in the form of text [Suhaili et al., 2021] and have grown rapidly more prominent in research since the year 2015.

A recent scoping review of chatbots in mental health revealed several pieces of interesting information within the field [Abd-Alrazaq et al., 2021]. Namely, the majority of chatbots focus on support for depression and autism, and controlled the conversation for therapy, training, and screening. The approach in this work is that of question-answering; that is, the goal of the model is to generalise online resources to provide answers that people may have about the included categories.

Bhagchandani and Nayak proposed the combination of two natural language processing models for a mental health chatbot framework [Bhagchandani and Nayak, 2022]. In this study, the authors first perform text classification using sentiment analysis to discern whether the user should be directed to a chatbot for a generic chat or another for therapy-based conversation. A similar approach was proposed in CareBot [Crasto et al., 2021], where conversational data was used along with the PHQ-9 and WHO-5 screening questionnaires to train a chatbot using a multimodal approach. The study recorded lower perplexity values for transformers compared to recurrent methods, but experimental observations revealed that 63% of the participants preferred the response generated by the Transformer over 22% for Long Short Term Memory (LSTM) networks and 15% for the Recurrent Neural Network (RNN).

In 2021, Deshpande and Warren proposed an additional module for a mental health chatbot which could detect users at risk of self-harm [Deshpande and Warren, 2021]; In their study, text classification experiments noted that the Bidirectional Encoder Representations from Transformers (BERT) could achieve 97% accuracy in recognising the risk within scraped Reddit data that were not part of the training dataset. BERT representations were also applied in a recent work, which found that it was a promising approach compared to classical approaches for the detection of mental health status from Reddit posts [Jiang et al., 2020]. Alongside the use of attention, several other methods have also been proposed to improve chatbots. These include data augmentation by paraphrasing [Bird et al., 2021, Joglekar, 2022], transfer learning [Prakash et al., 2020, Syed et al., 2021], reinforcement learning [Cuayáhuitl et al., 2019, Liu et al., 2020], and ensemble learning [Almansor et al., 2021, Bali et al., 2019].

Transformers are a new type of neural network that have recently seen a rapid rise in popularity, achieving state of the art performance in natural language processing, image captioning, image synthesis, classification, and audio processing [Lin et al., 2022]. Most relevant to this study are the studies exploring how transformer models achieve the current best performance metrics for the synthesis of text and answering of questions [Devlin and Chang, 2018, Lukovnikov et al., 2019, Radford et al., 2019, Shao et al., 2019]. According to the original paper [Vaswani et al., 2017], the attention values are calculated as the scaled dot product; Weights are calculated for each token within the input text as follows:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

where Q is the query token, an embedded representation of a word within a sequence. K represents keys, vectors of the sequence of tokens presented to the model, and V are values that are calculated when querying keys. In this study, Q, K, and V are from the same data source, and therefore the operation is described as self-attention. Each block also contains several attention heads, and thus the approach that this study implements is known as multi-headed self-attention (*MH*). This is simply calculated via the concatentation of h_i attention heads as follows:

$$MH(Q,K,V) = Concatenate(h_1, ..., h_h)W^O \tag{2}$$

The application of multi-headed attention has shown a significant improvement in ability compared to the conventional approach It is suggested that a shallower, wider model is more stable during the training process.

Fig. 1 shows a diagram of how the model uses embeddings as input and output, with a tokeniser used to transform both strings into encodings and vice versa.

3 METHOD

Within this section, the proposed methodology will be discussed, followed by work on optimisation of chatbots to answer mental health questions. The general approach of this work can be observed in Fig. 2; this section details each step of this process.

Initially, data from various sources were collected to form a large dataset. No single modern dataset is viable for large neural language models given their data requirements for effective generalisation [Sezgin et al., 2022]. Due to this, data from CounselChat[1], the Brain & Behaviour Research Foundation[2], the NHS[3,4], Wellness in Mind[5] and White Swan Foundation[6] were selected. Questions and answers are extracted, and questions are manually generated dependent on the information available, e.g. for the NHS definition of depression, questions such as "what is depression?" are imputed.

For preprocessing, all texts were converted to lowercase, and punctuation was removed in order to reduce the learning of irrelevant tokens. For example, the tokens "Hello", "hello", "Hello!", and "hello?" would all be treated as separate learnable tokens prior to this step. Then the vocabulary was limited to the most common 30,000 tokens to remove uncommon occurences that cannot be generalised. Following these steps, queries and answers are then denoted in the dataset with markup tags **<Q> ... </Q>** and **<A> ... </A>**, which are useful for several purposes: (i) to condition the model on separate types of text, (ii) to present the model with queries,

[1]Available online: https://counselchat.com [Last Accessed: 09/05/2023]
[2]Available online: https://www.bbrfoundation.org/faq/frequently-asked-questions-about-depression [Last Accessed: 09/05/2023]
[3]Available online: https://www.nhs.uk/mental-health/conditions/clinical-depression [Last Accessed: 09/05/2023]
[4]Available online: https://www.nhs.uk/mental-health/conditions/generalised-anxiety-disorder/overview [Last Accessed: 09/05/2023]
[5]Available online: https://www.wellnessinmind.org/frequently-asked-questions/ [Last Accessed: 09/05/2023]
[6]Available online: https://www.whiteswanfoundation.org/mental-health-matters/understanding-mental-health/mental-illness-faqs [Last Accessed: 09/05/2023]

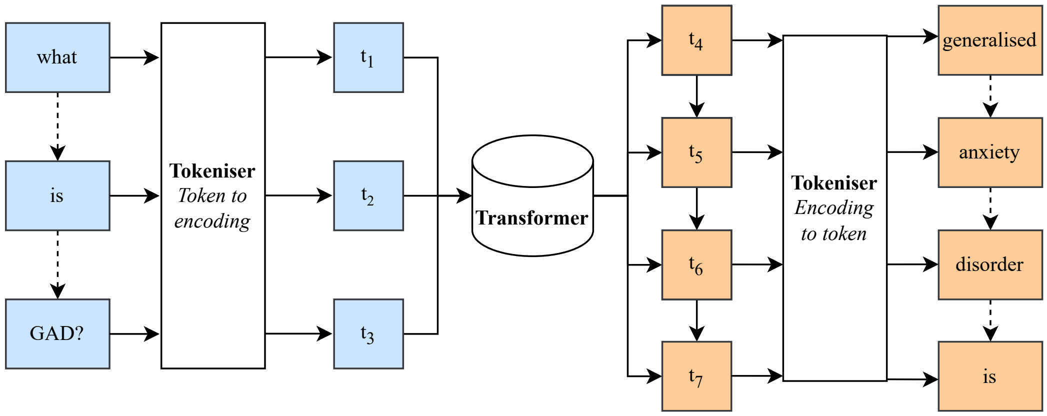

Figure 1: Diagram showing the use of a tokeniser to transform the text. Inputs are encoded and used for inference, encodings are output which are then transformed back into readable strings.

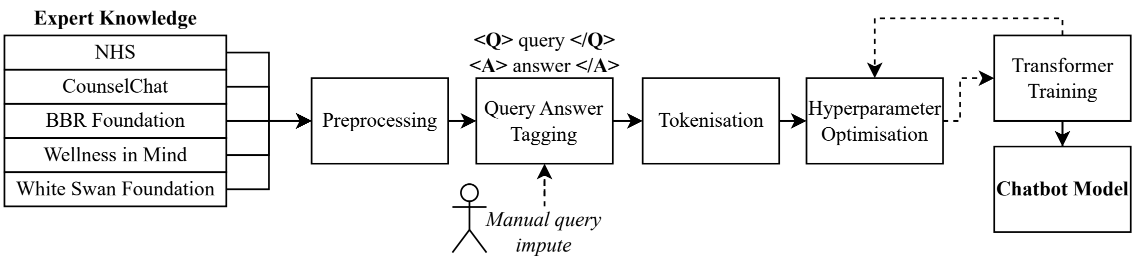

Figure 2: General diagram for the data preprocessing and training process for an optimised conversational chatbot model.

and (iii) to aid in the logic of ending the prediction loop when an answer has been generated.

With regards to the preprocessed data, a batch search of model hyperparameters were implemented for the generative transformer model. Starting from a random weight distribution, topologies of $\{2, 4, 8, 16\}$ attention heads was engineered and attached to one layer of $\{64, 128, 256, 512\}$ rectified linear units. Shallow networks are produced due to the data requirements of deeper models; although alarge dataset was collected, it is relatively close to the minimum requirements of a model following this learning method. In future, given more data, deeper networks could be explored. Models are trained and compared based on the validation metrics of accuracy and loss, with consideration also given to top-k accuracy where $k = 5$ and $k = 10$. Top-k metrics are important for deeper comparison of similarly-performing models, since it is a further measure of *how incorrect* a wrong prediction is. For example, two models selecting the correct token half of the time will both score 50% accuracy, but one model's second choice may more often be correct, suggesting that it is on a better track to generalise the data compared to the other.

To conclude the methodology shown in Fig. 2, a general diagram for the process of interfacing with the chatbot and inferring a response from the input query is shown in Fig. 3.

Table 1: Loss values for the transformer topology tuning experiments.

Dense Neurons	Attention Heads			
	2	***4***	***8***	***16***
64	0.64	0.56	**0.47**	0.91
128	0.65	0.58	**0.47**	1.16
256	0.65	0.59	0.48	1.37
512	0.64	0.59	1.42	1.72

4 RESULTS AND OBSERVATIONS

In this section, the observed metrics during the topology engineering for the transformer-based chatbots are presented before exploring some examples of its usage after training.

Table 1 and Table 2 show the loss and accuracy metrics for the 16 individual experiments, respectively. Two equally scoring models outperformed all others, which were eight attention heads succeeded by either 64 or 128 rectified linear units. Both of these models could predict the next token 88.65% of the time. Further to loss and accuracy metrics, Tables 3 and 4 show the top-k accuracy for $k = 5$ and $k = 10$, respectively. Beyond the initial results, these tables show us that the option of using 128 neurons in the layer prior to token prediction gives a slightly higher result. These were 96.49% (against 96.41%) and 97.88% (against 97.82%). The 8-headed,

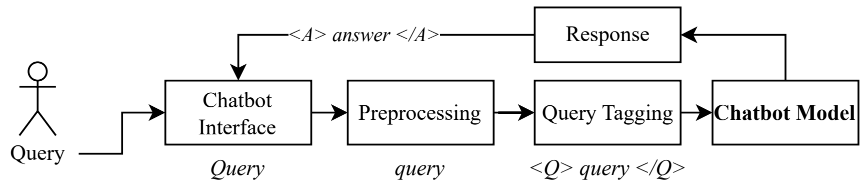

Figure 3: Diagram of the inference process for the trained chatbot model interface.

Table 2: Accuracy values for the transformer topology tuning experiments.

Dense Neurons	Attention Heads			
	2	***4***	***8***	***16***
64	84.13	86.23	**88.65**	79.5
128	83.88	85.95	**88.65**	74.3
256	83.81	85.46	88.02	69.98
512	84.01	85.44	68.9	63.86

Table 3: Top-5 accuracy values for the transformer topology tuning experiments.

Dense Neurons	Attention Heads			
	2	***4***	***8***	***16***
64	94.86	95.73	96.41	90.89
128	94.84	95.53	**96.49**	87.15
256	94.78	95.43	96.34	84.26
512	94.85	95.31	87.7	78.79

Table 4: Top-10 accuracy values for the transformer topology tuning experiments.

Dense Neurons	Attention Heads			
	2	***4***	***8***	***16***
64	96.92	97.64	97.82	93.92
128	96.94	97.36	**97.88**	90.92
256	96.83	97.29	97.77	88.78
512	96.95	97.23	88.48	84.12

128-neuron model is therefore selected as the best candidate for further exploration.

To provide examples of interaction with the trained model with *<Q>* queries provided by a human and *<A>* answers output by the model, the following is an excerpt from a conversation about depression[7]:

<Q> what is depression **</Q>**
<A> clinical depression is a serious condition that negatively affects how a person thinks feels and behaves **</A>**
<Q> what are the symptoms of depression **</Q>**
<A> the symptoms of depression can be complex and vary widely between people if youre depressed you may feel sad hopeless and lose interest in things you used to enjoy **</A>**
<Q> how is depression diagnosed **</Q>**
<A> the first step to being diagnosed is to visit a doctor for a medical evaluation certain medications and some medical conditions such as thyroid disorder can cause similar symptoms as depression **</A>**
<Q> are certain people more likely to get depressed than others **</Q>**
<A> you may be more vulnerable to depression if you have certain personality traits such as low selfesteem or being overly selfcritical this may be because of the genes youve inherited from your parents your early life experiences or both **</A>**
<Q> can you tell me about depression after childbirth **</Q>**
<A> sometimes new mothers fathers or partners develop depression after they have a baby this is known as postnatal depression and its treated in a similar way to other types of depression with talking therapies and antidepressant medicines **</A>**

In addition, the following is an excerpt from a conversation about anxiety:

<Q> is anxiety normal **</Q>**
<A> everyone has feelings of anxiety at some point in their life for example you may feel worried and anxious about sitting an exam or having a medical test or job interview during times like these feeling anxious can be perfectly normal **</A>**
<Q> if i feel too anxious should i see a doctor **</Q>**
<A> although feelings of anxiety at certain times are completely normal you should see a gp if anxiety is affecting your daily life or causing you distress **</A>**
<Q> what are the symptoms of anxiety **</Q>**
<A> gad can cause both psychological mental and physical symptoms these vary from person to person but can include feeling restless or worried having trouble concentrating or sleeping dizziness or heart palpitations **</A>**
<Q> what does gad mean **</Q>**

[7] Note: all grammar has been removed from the training data to enable generalisation of tokens with identical meaning

<A> gad stands for general anxiety disorder **</A>**
<Q> is there any way i can feel less anxious **</Q>**
<A> there are many things that can do yourself to help reduce your anxiety such as going on a selfhelp course exercising regularly stopping smoking looking after your physical health **</A>**

As can be observed from the aforementioned conversations, interaction with the most optimal model leads to examples where queries can be effectively answered and advice given following training from the verified sources. Terms such as GAD (*General Anxiety Disorder*) are more likely to appear in the outputs since they were abbreviated more often than not within the training data; in this case, it was possible to ask the chatbot to clarify this term. Reducing the number of unique tokens via removing grammar aided in training with a dataset of this given size, but results in none being output. In future, more natural conversation would be enabled through either learning from a grammatically-correct dataset, or correcting the chatbot output prior to the response being printed to an interface.

5 CONCLUSION AND FUTURE WORK

In this work, the engineering and applications of transformer-based chatbots are explored to answer questions with a focus on mental health support. Specifically, the focus is on queries surrounding depression and anxiety from respected and verified sources. To conclude this work, chatbots have the potential to play a significant role in supporting people suffering from mental health stigma. The use of attention mechanism techniques to build chatbots from transformers, which are large language models, seem to lead to the creation of engaging conversational systems. The results of this study demonstrate the potential of chatbots to provide easily accessible and anonymous support to people who may otherwise be discouraged from seeking help due to stigma. However, with these findings considered, it is also important to acknowledge the limitations and challenges of using chatbots for mental health support. More research from medical and psychological backgrounds is needed to fully understand the limitations of chatbots and ensure that they are developed and deployed ethically and responsibly.

Alongside future work regarding ethics, there are also limitations to this study that should be explored. Firstly, data availability is a concern; although we collected a large dataset for this study, this laborious process led to only the minimal amount of data to train such models. In the future, more data could be collected and experiments could be reimplemented to further generalisation. Additionally, methods such as transfer learning and data augmentation could be explored as alternatives to alleviate this limitation. To engineer the topologies, we performed a batch search; this could be further improved through metaheuristic hyperparameter optimisation to automate this process. Although this would likely lead to a better model, it would require far more computational resources and time.

In addition to future experiments, examples such as the chatbot outputting "GAD" (instead of *General Anxiety Disorder*) show how the model can be affected when the majority of terms are abbreviated within the training data. In the future, the application may be more informative if abbreviations are replaced with definitions as an added data preprocessing step.

Finally, in conclusion, this study highlights the importance of continuing research and development in the field of mental health technology. By exploring the potential of chatbots to provide support to individuals experiencing depression and anxiety, we can work toward creating innovative and effective solutions to promote mental well-being.

REFERENCES

Alaa A Abd-Alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M Bewick, Peter Gardner, and Mowafa Househ. 2019. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics* 132 (2019), 103978.

Alaa A Abd-Alrazaq, Mohannad Alajlani, Nashva Ali, Kerstin Denecke, Bridgette M Bewick, and Mowafa Househ. 2021. Perceptions and opinions of patients about mental health chatbots: scoping review. *Journal of medical Internet research* 23, 1 (2021), e17828.

Ebtesam Hussain Almansor, Farookh Khadeer Hussain, and Omar Khadeer Hussain. 2021. Supervised ensemble sentiment-based framework to measure chatbot quality of services. *Computing* 103 (2021), 491–507.

Manish Bali, Samahit Mohanty, Subarna Chatterjee, Manash Sarma, and Rajesh Puravankara. 2019. Diabot: a predictive medical chatbot using ensemble learning. *International Journal of Recent Technology and Engineering* 8, 2 (2019), 6334–6340.

Himanshu Bansal and Rizwan Khan. 2018. A review paper on human computer interaction. *Int. J. Adv. Res. Comput. Sci. Softw. Eng* 8, 4 (2018), 53.

Anushka Bhagchandani and Aryan Nayak. 2022. Deep Learning Based Chatbot Framework for Mental Health Therapy. In *Advances in Data and Information Sciences: Proceedings of ICDIS 2021*. Springer, 271–281.

Jordan J Bird, Anikó Ekárt, and Diego R Faria. 2021. Chatbot Interaction with Artificial Intelligence: human data augmentation with T5 and language transformer ensemble for text classification. *Journal of Ambient Intelligence and Humanized Computing* (2021), 1–16.

Reuben Crasto, Lance Dias, Dominic Miranda, and Deepali Kayande. 2021. CareBot: A Mental Health ChatBot. In *2021 2nd International Conference for Emerging Technology (INCET)*. IEEE, 1–5.

Heriberto Cuayáhuitl, Donghyeon Lee, Seonghan Ryu, Yongjin Cho, Sungja Choi, Satish Indurthi, Seunghak Yu, Hyungtak Choi, Inchul Hwang, and Jihie Kim. 2019. Ensemble-based deep reinforcement learning for chatbots. *Neurocomputing* 366 (2019), 118–130.

Saahil Deshpande and Jim Warren. 2021. Self-Harm Detection for Mental Health Chatbots.. In *MIE*. 48–52.

Jacob Devlin and Ming-Wei Chang. 2018. Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing. *Google AI Blog. Weblog.[Online] Available from: https://ai. googleblog. com/2018/11/open-sourcing-bertstate-of-art-pre. html [Accessed 4 December 2019]* (2018).

Richard G Frank and Sherry A Glied. 2006. *Better but not well: Mental health policy in the United States since 1950*. JHU Press.

Terry Hanley and Claire Wyatt. 2021. A systematic review of higher education students' experiences of engaging with online therapy. *Counselling and Psychotherapy Research* 21, 3 (2021), 522–534.

Zheng Ping Jiang, Sarah Ita Levitan, Jonathan Zomick, and Julia Hirschberg. 2020. Detection of mental health from reddit via deep contextualized representations. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*. 147–156.

Chaitanya Joglekar. 2022. *WOzBot: A Wizard of Oz Based Method for Chatbot Response Improvement*. Master's thesis. Trinity College Dublin.

Edgar Jones and Simon Wessely. 2005. *Shell shock to PTSD: Military psychiatry from 1900 to the Gulf War*. Psychology Press.

Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. 2022. A survey of transformers. *AI Open* (2022).

Jianfeng Liu, Feiyang Pan, and Ling Luo. 2020. Gochat: Goal-oriented chatbots with hierarchical reinforcement learning. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1793–1796.

Denis Lukovnikov, Asja Fischer, and Jens Lehmann. 2019. Pretrained transformers for simple question answering over knowledge graphs. In *International Semantic Web Conference*. Springer, 470–486.

Kolla Bhanu Prakash, Y Nagapawan, N Lakshmi Kalyani, and V Pradeep Kumar. 2020. Chatterbot implementation using transfer learning and LSTM encoder-decoder architecture. *International Journal* 8, 5 (2020).

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. (2019).

Emre Sezgin, Joseph Sirrianni, and Simon L Linwood. 2022. Operationalizing and implementing pretrained, large artificial intelligence linguistic models in the US health care system: outlook of generative pretrained transformer 3 (GPT-3) as a service model. *JMIR medical informatics* 10, 2 (2022), e32875.

Taihua Shao, Yupu Guo, Honghui Chen, and Zepeng Hao. 2019. Transformer-based neural network for answer selection in question answering. *IEEE Access* 7 (2019), 26146–26156.

Amy E Sickel, Jason D Seacat, and Nina A Nabors. 2014. Mental health stigma update: A review of consequences. *Advances in Mental Health* 12, 3 (2014), 202–215.

Sinarwati Mohamad Suhaili, Naomie Salim, and Mohamad Nazim Jambli. 2021. Service chatbots: A systematic review. *Expert Systems with Applications* 184 (2021), 115461.

Zeeshan Haque Syed, Asma Trabelsi, Emmanuel Helbert, Vincent Bailleau, and Christian Muths. 2021. Question answering chatbot for troubleshooting queries based on transfer learning. *Procedia Computer Science* 192 (2021), 941–950.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

WHO. 2021. Depression. https://www.who.int/news-room/fact-sheets/detail/depression